

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
28 July 2005 (28.07.2005)

PCT

(10) International Publication Number
WO 2005/068089 A2

(51) International Patent Classification⁷: **B06F**

(21) International Application Number: **PCT/GB2005/000106**

(22) International Filing Date: 14 January 2005 (14.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0400974.2 16 January 2004 (16.01.2004) GB

(71) Applicant (*for all designated States except US*): SOLEXA LIMITED [GB/GB]; Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB).

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): COX, Anthony, James [GB/GB]; Solexa Limited, Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB).

(74) Agents: PALMER, Jonathan, Richard et al.; Boult Wade Tennant, Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/068089 A2

(54) Title: MULTIPLE INEXACT MATCHING

(57) Abstract: Methods and apparatus are provided for seeking, in a target sequence, an exact or inexact match to a query sequence, or each of a large number of query sequences, which could be DNA fragments for which an approximate match is sought in a genome. Initially, each query sequence is divided into segments, which are distributed between first and second segment groups. An efficient algorithm is then used to seek potential matches between a first group of segments of each query sequence and a corresponding first group of segments of each target sequence fragment.